

MacWilliams Identities for Terminated Convolutional Codes

G. David Forney, Jr.

Laboratory for Information and Decision Systems
 Massachusetts Institute of Technology
 Cambridge, MA 02139
 Email: forneyd@comcast.net

Abstract—Shearer and McEliece [8] showed that there is no MacWilliams identity for the free distance spectra of orthogonal linear convolutional codes. We show that on the other hand there does exist a MacWilliams identity between the generating functions of the weight distributions per unit time of a linear convolutional code \mathcal{C} and its orthogonal code \mathcal{C}^\perp , and that this distribution is as useful as the free distance spectrum for estimating code performance. These observations are similar to those made recently by Bocharova *et al.* [1]; however, we focus on terminating by tail-biting rather than by truncation.

I. INTRODUCTION

Finding a MacWilliams-type identity for convolutional codes is a problem of long standing [8]. For a linear time-invariant convolutional code \mathcal{C} over a finite field, the most commonly studied distance distribution is the *free (Hamming) distance spectrum*, namely, the distribution of (Hamming) weights of codewords in \mathcal{C} that start and end in the zero state without passing through an intermediate zero state. Shearer and McEliece [8] showed by example that the free distance spectrum of \mathcal{C} does not in general determine that of \mathcal{C}^\perp , and therefore that there could be no MacWilliams identity for such distributions.

Gluesing-Luerssen and Schneider (GLS) have recently formulated [6] and proved [7] a MacWilliams-type identity for convolutional codes involving the Hamming weight adjacency matrix (HWAM) of a convolutional code \mathcal{C} and the HWAM of its orthogonal code \mathcal{C}^\perp . In [3], [4], the GLS result was proved in a different way, and generalized to various kinds of weight adjacency matrices and to group codes defined on graphs.

More recently, Bocharova, Hug, Johannesson and Kudryashov [1] have proved a different MacWilliams-type identity for truncations of a convolutional code \mathcal{C} and its orthogonal code \mathcal{C}^\perp , and have shown that by letting the truncation length become large, an approximation to the free distance spectrum can be obtained. In this paper, which is mostly based on [4], we derive similar results for weight distributions of block codes obtained by various kinds of termination procedures, of which we regard tail-biting as the nicest. We argue that these alternative distributions are just as useful for estimating code performance as the free distance spectrum. These results effectively answer the original question posed by Shearer and McEliece [8].

II. TERMINATED CONVOLUTIONAL CODES

A general method for approximating the free distance spectrum of a linear convolutional code \mathcal{C} is to derive a series of block codes \mathcal{C}_N of length N from \mathcal{C} by some sort of termination procedure, and then to study the distance distributions of \mathcal{C}_N as $N \rightarrow \infty$. As we shall see in Section III, for any of the termination methods below, the distance distribution of \mathcal{C}_N , normalized by N , approaches the free distance spectrum of \mathcal{C} for $d_{\text{free}} \leq d < 2d_{\text{free}}$. However, the usual termination methods are problematic if we are also interested in the distance distribution of the orthogonal convolutional code \mathcal{C}^\perp .

For example, the most common termination method is to take the *subcode* $\mathcal{C}_{[0,N]}$ of \mathcal{C} , consisting of all sequences in \mathcal{C} whose support is contained in the interval $[0, N] = \{k \in \mathbb{Z} \mid 0 \leq k < N\}$ (*i.e.*, all code sequences that pass through the zero state at times 0 and N , restricted to $[0, N]$), which is effectively a block code of length N time units. As a subcode of \mathcal{C} , $\mathcal{C}_{[0,N]}$ has at least the minimum (free) distance of \mathcal{C} . However, the orthogonal code to $\mathcal{C}_{[0,N]}$ is the *projection* $(\mathcal{C}^\perp)_{|[0,N]}$ of the orthogonal convolutional code \mathcal{C}^\perp onto the interval $[0, N]$ (*i.e.*, all orthogonal code sequences that pass through any state at times 0 and N , restricted to $[0, N]$). In general, a projection $(\mathcal{C}^\perp)_{|[0,N]}$ has low-weight codewords, no matter how large N becomes.

Bocharova *et al.* [1] have considered another kind of terminated code that they call a *truncated code*, which we will denote by $\mathcal{C}_{\triangleleft[0,N]}$. Such a code may be described as the subcode $\mathcal{C}_{[0,\infty)}$ projected onto $[0, N]$: *i.e.*,

$$\mathcal{C}_{\triangleleft[0,N]} = (\mathcal{C}_{[0,\infty)})_{|[0,N]}.$$

By projection/subcode duality, the dual code $(\mathcal{C}_{\triangleleft[0,N]})^\perp$ is

$$((\mathcal{C}_{[0,\infty)})_{|[0,N]})^\perp = ((\mathcal{C}_{[0,\infty)})^\perp)_{|[0,N]} = ((\mathcal{C}^\perp)_{|[0,\infty)})_{|[0,N]};$$

i.e., the subcode defined on $[0, N]$ of the projection of the orthogonal convolutional code \mathcal{C}^\perp onto $[0, \infty)$. Alternatively, $(\mathcal{C}_{\triangleleft[0,N]})^\perp$ may be defined as a reverse-truncated code [1]

$$(\mathcal{C}^\perp)_{\triangleright[0,N]} = ((\mathcal{C}^\perp)_{(-\infty,N)})_{|[0,N]}.$$

Thus there is a MacWilliams identity between the weight distributions of $\mathcal{C}_{\triangleleft[0,N]}$ and $(\mathcal{C}^\perp)_{\triangleright[0,N]}$. Moreover, unlike the

subcode $\mathcal{C}_{[0,N]}$ or the projection $\mathcal{C}_{|[0,N]}$, a truncated code has the same rate as \mathcal{C} . However, truncated codes have small minimum distance.

A more elegant method of terminating a convolutional code \mathcal{C} is via tail-biting. The *tail-biting terminated code* $\mathcal{C}_{||[0,N]}$ of the convolutional code \mathcal{C} on the interval $[0, N]$ is the set of all codewords in \mathcal{C} that pass through the same state at times 0 and N , restricted to $[0, N]$. For large enough N , the tail-biting code $\mathcal{C}_{||[0,N]}$ has the same minimum distance as \mathcal{C} ; moreover, $\mathcal{C}_{||[0,N]}$ has the same rate as \mathcal{C} . Most importantly, the orthogonal code to $\mathcal{C}_{||[0,N]}$ is $(\mathcal{C}^\perp)_{||[0,N]}$, the tail-biting terminated code of \mathcal{C}^\perp on the same interval [2]. Thus there is a MacWilliams identity between the weight distributions of $\mathcal{C}_{||[0,N]}$ and $(\mathcal{C}^\perp)_{||[0,N]}$. Finally, we will see that these distributions approach the free distance spectra of \mathcal{C} and \mathcal{C}^\perp nicely as $N \rightarrow \infty$.

Example 1 (rate-1/2 4-state binary linear convolutional code). Consider the rate-1/2 binary linear time-invariant convolutional code \mathcal{C} with degree-2 generators $(1 + D^2, 1 + D + D^2)$, in standard D -transform notation. A minimal encoder for this code is the linear time-invariant system with impulse response $(11, 01, 11, 00, \dots)$, which has the 4-state trellis section shown in Figure 1(a). The orthogonal convolutional code \mathcal{C}^\perp is the rate-1/2 binary linear convolutional code with generators $(1 + D + D^2, 1 + D^2)$, which has a minimal 4-state linear encoder with impulse response $(11, 10, 11, 00, \dots)$, and the 4-state trellis section shown in Figure 1(b).

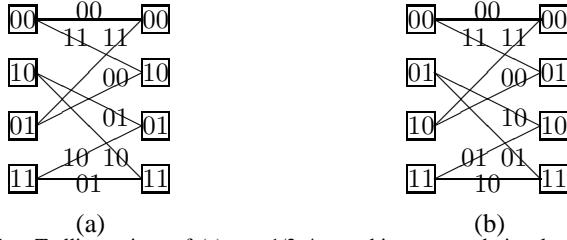


Fig. 1. Trellis sections of (a) rate-1/2 4-state binary convolutional code \mathcal{C} ; (b) orthogonal code \mathcal{C}^\perp .

We now consider various methods of terminating this convolutional code \mathcal{C} with a block length of $N = 4$. The subcode $\mathcal{C}_{[0,4]}$ is the (8, 2) binary linear block code generated by the two generators

$$\begin{array}{cccc} 11 & 01 & 11 & 00 \\ 00 & 11 & 01 & 11 \end{array}$$

The minimum distance of this block code is the same as that of \mathcal{C} , namely 5, although its rate is lower. The orthogonal code to the subcode $\mathcal{C}_{[0,4]}$ is the projection $(\mathcal{C}^\perp)_{|[0,4]}$ of the orthogonal convolutional code \mathcal{C}^\perp , which is the (8, 6) binary linear block code generated by the six generators

$$\begin{array}{cccc} 11 & 00 & 00 & 00 \\ 10 & 11 & 00 & 00 \\ 11 & 10 & 11 & 00 \\ 00 & 11 & 10 & 11 \\ 00 & 00 & 11 & 10 \\ 00 & 00 & 00 & 11 \end{array}$$

The minimum distance of this block code is 2, less than that of \mathcal{C}^\perp , although its rate is higher.

The truncated code $\mathcal{C}_{\triangleleft[0,4]}$ is the (8, 4) binary linear block code generated by

$$\begin{array}{cccc} 11 & 01 & 11 & 00 \\ 00 & 11 & 01 & 11 \\ 00 & 00 & 11 & 01 \\ 00 & 00 & 00 & 11 \end{array}$$

The minimum distance of this block code is 2, but its rate is the same as that of \mathcal{C} . Its orthogonal code $(\mathcal{C}^\perp)_{\triangleright[0,4]}$ is the (8, 4) binary linear block code generated by

$$\begin{array}{cccc} 11 & 00 & 00 & 00 \\ 10 & 11 & 00 & 00 \\ 11 & 10 & 11 & 00 \\ 00 & 11 & 10 & 11 \end{array}$$

which has the same parameters.

The tail-biting terminated code $\mathcal{C}_{||[0,4]}$ is the (8, 4) binary linear block code generated by

$$\begin{array}{cccc} 11 & 01 & 11 & 00 \\ 00 & 11 & 01 & 11 \\ 11 & 00 & 11 & 01 \\ 01 & 11 & 00 & 11 \end{array}$$

whereas the orthogonal tail-biting terminated code $(\mathcal{C}^\perp)_{||[0,N]}$ is the (8, 4) binary linear block code generated by the four generators

$$\begin{array}{cccc} 11 & 10 & 11 & 00 \\ 00 & 11 & 10 & 11 \\ 11 & 00 & 11 & 10 \\ 10 & 11 & 00 & 11 \end{array}$$

Both of these codes have a minimum distance of only 2 (e.g., for paths such as 01 00 01 00 from state 10 to state 10). However, for $N \geq 10$, it turns out that the minimum distance of both tail-biting terminated codes is 5, the same as the minimum distance of \mathcal{C} or \mathcal{C}^\perp . \square

III. FREE DISTANCE SPECTRA FOR CONVOLUTIONAL CODES FROM TERMINATED CODES

Let us now consider how the free distance spectrum of a linear time-invariant convolutional code \mathcal{C} may be derived from the weight distribution of a terminated code of length N as $N \rightarrow \infty$.

Without loss of generality, we may assume that \mathcal{C} is generated by a minimal encoder, which is necessarily *non-catastrophic*: i.e., the unique state sequence associated with the all-zero code sequence is the all-zero state sequence. Consequently, the lowest-weight words of a terminated code as $N \rightarrow \infty$ must be those that pass through the zero state almost all of the time. These code sequences are as follows, for the various termination methods we have considered:

- If we terminate to the subcode $\mathcal{C}_{[0,N]}$, then code sequences start and end in the zero state, and the lowest-weight sequences correspond to the lowest-weight sequences in the free distance spectrum. If the minimum

free distance is d_{free} , then for $d_{\text{free}} \leq d < 2d_{\text{free}}$ there will be approximately $N \times N_d$ sequences in the terminated code of weight d , where N_d is the number of code sequences of weight d in the free distance spectrum of \mathcal{C} . Thus, for $d_{\text{free}} \leq d < 2d_{\text{free}}$, the weight distribution per unit time of \mathcal{C} is the limit of the weight distribution of $\mathcal{C}_{[0,N]}$ normalized by (divided by) N as $N \rightarrow \infty$. For $d \geq 2d_{\text{free}}$, there will be overcounting—e.g., two sequences of weight d_{free} may be counted as one of weight $2d_{\text{free}}$ —but we will argue below that such overcounting should not affect estimates of code performance.

- If we terminate to the projection $\mathcal{C}_{|[0,N]}$, then code sequences can start and end in any state, and there will be low-weight sequences starting with a low-weight state transition $s \rightarrow 0$, remaining in state 0 for nearly N time units, and then ending with a low-weight transition $0 \rightarrow s'$, where s and s' are not both 0. Thus the minimum distance of $\mathcal{C}_{|[0,N]}$ will be less than d_{free} for all N . However, the number of such low-weight sequences remains constant, so after normalization we will eventually see the same normalized weight distribution as for $\mathcal{C}_{[0,N]}$.
- If we terminate to the truncated code $\mathcal{C}_{\triangleleft[0,N]}$, then by the same argument we will eventually see the same normalized weight distribution. In this case the total weight of a code sequence starting in the zero state, remaining there for nearly N time units, and then ending with a low-weight transition $0 \rightarrow s$, is only that of the low-weight transition $0 \rightarrow s$. However, again the number of such low-weight sequences remains constant, so after normalization we will eventually see the correct normalized weight distribution.
- If we terminate to the tail-biting code $\mathcal{C}_{||[0,N]}$, then by the same argument we will eventually see the same normalized weight distribution. Note however that in this case the total weight of a code sequence starting with a low-weight transition $s \rightarrow 0$, remaining in the zero state for nearly N time units, and then ending with a low-weight transition $0 \rightarrow s$, must be at least d_{free} , since the ending sequence (corresponding to the state transition $0 \rightarrow s$) followed by the starting sequence (corresponding to $s \rightarrow 0$) must be a code sequence. Thus the minimum distance of $\mathcal{C}_{||[0,N]}$ must equal d_{free} for large enough N .

We conclude that as $N \rightarrow \infty$ the normalized weight distribution of any of these terminated codes approaches the free distance spectrum of \mathcal{C} for $d_{\text{free}} \leq d < 2d_{\text{free}}$. However, only the tail-biting termination has the same rate as \mathcal{C} and, for N large enough, the same minimum distance d_{free} .

Finally, we argue that the normalized weight distribution of any of these terminated codes \mathcal{C}_N must yield the same estimate of code performance over N time units as the free distance spectrum of \mathcal{C} , if these estimates are accurate. The probability of error event $P(\mathcal{E})$ of \mathcal{C} per unit time may be estimated using the free distance spectrum. The probability of any error in N time units is then estimated as $NP(\mathcal{E})$. If this is a good

estimate (implying $N < 1/P(\mathcal{E})$), then the probability of two or more error events in N time units must be negligible. But the probability of any error in decoding \mathcal{C} over N time units is essentially the same as the probability of block decoding error in decoding \mathcal{C}_N , which may be estimated by the weight distribution of N , which counts codewords that include two or more error events. If the probability of two or more error events in N time units is negligible, then an estimate based on the weight distribution of \mathcal{C}_N must approximately agree with an estimate based on the free distance spectrum of \mathcal{C} .

IV. WEIGHT ADJACENCY MATRICES AND WEIGHT GENERATING FUNCTIONS

We now show how Hamming weight generating functions for terminations of a linear time-invariant convolutional code \mathcal{C} may be derived from the Hamming weight adjacency matrix of a minimal linear time-invariant encoder for \mathcal{C} . This will allow us to state MacWilliams identities for terminated convolutional codes, and to estimate code performance.

Given a linear time-invariant encoder for a convolutional code \mathcal{C} with state space \mathcal{S} and symbol alphabet \mathcal{A} , the Hamming weight adjacency matrix (HWAM) is the matrix $\Lambda(x)$ indexed by $\mathcal{S} \times \mathcal{S}$ whose elements are

$$\Lambda_{ss'}(x) = \sum_{a \in \mathcal{T}(s, s')} x^{w(a)},$$

where x is an indeterminate, $\mathcal{T}(s, s')$ is the subset of symbols $a \in \mathcal{A}$ such that the state/symbol transition (“branch”) $(s, a, s') \in \mathcal{S} \times \mathcal{A} \times \mathcal{S}$ actually occurs in the encoder, and $w(a)$ is the Hamming weight of the symbol $a \in \mathcal{A}$.

Example 1 (cont.). For the rate-1/2 binary convolutional code \mathcal{C} of Example 1, the HWAM of the encoder of Figure 1(a) is

$$\Lambda(x) = \begin{array}{c|ccccc} s/s' & 00 & 10 & 01 & 11 \\ \hline 00 & 1 & x^2 & 0 & 0 \\ 10 & 0 & 0 & x & x \\ 01 & x^2 & 1 & 0 & 0 \\ 11 & 0 & 0 & x & x \end{array}$$

For the orthogonal code \mathcal{C}^\perp , the HWAM of the encoder of Figure 1(b) is

$$\hat{\Lambda}(x) = \begin{array}{c|ccccc} s/s' & 00 & 10 & 01 & 11 \\ \hline 00 & 1 & 0 & x^2 & 0 \\ 10 & x^2 & 0 & 1 & 0 \\ 01 & 0 & x & 0 & x \\ 11 & 0 & x & 0 & x \end{array}$$

which in this case is the transpose of $\Lambda(x)$. \square

It is shown in [6], [7], [3], [4] how the HWAM $\Lambda(x)$ of a minimal encoder for \mathcal{C} determines the HWAM $\hat{\Lambda}(X)$ of a minimal encoder for \mathcal{C}^\perp and *vice versa* via a MacWilliams-type identity, but we will not need that result here.

Now it is easy to see that if we take N consecutive trellis sections of a minimal encoder for \mathcal{C} as a single section, then the HWAM of this length- N trellis section is simply the N th power $\Lambda^N(x)$ of the basic HWAM $\Lambda(x)$.

$$\Lambda^4(x) = \begin{bmatrix} 1 + 2x^5 + x^6 & x^2 + x^3 + x^4 + x^7 & x^3 + 2x^4 + x^5 & x^3 + 2x^4 + x^5 \\ x^3 + 2x^4 + x^5 & x^2 + x^3 + x^5 + x^6 & 2x^3 + x^4 + x^6 & 2x^3 + x^4 + x^6 \\ x^2 + x^3 + x^4 + x^7 & x^2 + x^3 + x^5 + x^6 & x^2 + x^3 + x^5 + x^6 & x^2 + x^3 + x^5 + x^6 \\ x^3 + 2x^4 + x^5 & x^2 + x^3 + x^5 + x^6 & 2x^3 + x^4 + x^6 & 2x^3 + x^4 + x^6 \end{bmatrix}.$$

Fig. 2. HWAM $\Lambda^4(x)$ of a section of $N = 4$ time units of Example 1 code \mathcal{C} .

Example 1 (cont.). Given the HWAM $\Lambda(x)$ above for a minimal encoder of our example code \mathcal{C} , the HWAM of a section consisting of $N = 2$ time units of this code is

$$\Lambda^2(x) = \begin{bmatrix} 1 & x^2 & x^3 & x^3 \\ x^3 & x & x^2 & x^2 \\ x^2 & x^4 & x & x \\ x^3 & x & x^2 & x^2 \end{bmatrix}.$$

This shows that there is exactly one path from each state at time k to each state at time $k + 2$, and that the minimum Hamming weight of any of these paths (other than the zero path) is 1.

For a section consisting of $N = 4$ time units of this code, the HWAM $\Lambda^4(x)$ is given in Fig. 2. This HWAM shows that there are four paths from each state at time k to each state at time $k + 4$, and that the minimum nonzero Hamming weight of any of these paths is 2. \square

The weight generating functions of various terminated codes of \mathcal{C} can now be read from these weight adjacency matrices. Since the subcode $\mathcal{C}_{[0,N]}$ is the set of all sequences in \mathcal{C} that pass through the zero states at times 0 and N , its weight generating function is simply the $(0, 0)$ element of $\Lambda^N(x)$. Similarly, since the projection $\mathcal{C}_{|[0,N]}$ is the set of all sequences in \mathcal{C} that pass through any states at times 0 and N , its weight generating function is the sum of all elements of $\Lambda^N(x)$.

Since the truncated code $\mathcal{C}_{\triangleleft[0,N]}$ is the set of all sequences in \mathcal{C} that pass through the zero state at times 0 and any state at time N , its weight generating function is the sum of all elements in the first row of $\Lambda^N(x)$. Similarly, the weight generating function of $\mathcal{C}_{\triangleright[0,N]}$ is the sum of all elements in the first column of $\Lambda^N(x)$.

Finally, since the tail-biting termination $\mathcal{C}_{||[0,N]}$ is the set of all sequences in \mathcal{C} that pass through the same states in S_0 and S_N , its weight generating function is the sum of all diagonal elements of $\Lambda^N(x)$; i.e., its trace $\text{Tr}(\Lambda^N(x))$. Since $\mathcal{C}_{||[0,N]}$ and $(\mathcal{C}^\perp)_{||[0,N]}$ are orthogonal block codes, there is a MacWilliams identity between their weight generating functions.

Example 1 (cont.). For the rate-1/2 binary convolutional code of Example 1, the Hamming weight generating function of the the tail-biting termination $\mathcal{C}_{||[0,4]}$ of length 4 is the trace of $\Lambda^4(x)$, namely $1 + 2x^2 + 4x^3 + x^4 + 4x^5 + 4x^6$. Since $\hat{\Lambda}^4(x)$ is the transpose of $\Lambda^4(x)$, the orthogonal tail-biting terminated code $(\mathcal{C}^\perp)_{||[0,4]}$ is an equivalent code with the same Hamming weight generating function. It is easy to check that this Hamming weight generating function is indeed invariant under the MacWilliams transform. \square

Using tail-biting terminated codes, and normalizing the weight distribution by dividing by N , we have that the generating function of the normalized Hamming weight distribution of \mathcal{C} is

$$g_{\mathcal{C}}(x) = \lim_{N \rightarrow \infty} \frac{1}{N} \text{Tr}(\Lambda^N(x)).$$

Moreover, there is a MacWilliams identity between $g_{\mathcal{C}}(x)$ and $g_{\mathcal{C}^\perp}(x)$. The performance of \mathcal{C} is determined by $g_{\mathcal{C}}(x)$, and that of \mathcal{C}^\perp by $g_{\mathcal{C}^\perp}(x)$. (Similar observations are made in [1], using truncated codes.)

Example 1 (cont.). For a section consisting of $N = 16$ time units of the rate-1/2 binary convolutional code \mathcal{C} of Example 1, the HWAM $\Lambda^{16}(x)$ (modulo x^8) is given in Fig. 3 at the top of the next page. Notice that

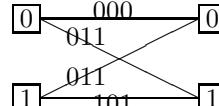
$$\text{Tr}(\Lambda^{16}(x)) = 1 + 16x^5 + 32x^6 + 64x^7 + \dots,$$

so that normalizing the distribution by dividing by $N = 16$ already gives the precise free distance spectrum of \mathcal{C} for $d < 8$, namely $x^5 + 2x^6 + 4x^7 + \dots$. Thus the convergence to the limiting generating function $g_{\mathcal{C}}(x)$ is rapid and exact. This property of tail-biting terminations is not shared by other kinds of terminations. \square

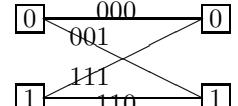
It appears that the behavior of $g_{\mathcal{C}}(x)$ might be analyzed by using an extension of Perron-Frobenius theory to generating function matrices, as in [5]; however, we have not attempted such an analysis.

Example 2 (cf. [8], [1]). The two codes proposed by Shearer and McEliece [8] for their counterexample make an excellent example. The first code is a rate-1/3 binary linear time-invariant convolutional code \mathcal{C}_1 generated by the degree-1 generators $(1, 1 + D, D)$, i.e., \mathcal{C}_1 is generated by a minimal encoder with impulse response $(110, 011, 000, \dots)$, whose trellis section is shown in Figure 4(a). The HWAM of this encoder is

$$\Lambda_1(x) = \begin{bmatrix} 1 & x^2 \\ x^2 & x^2 \end{bmatrix}.$$



(a)



(b)

Fig. 4. Trellis sections of (a) rate-1/3 2-state binary convolutional code \mathcal{C}_1 ; (b) similar code \mathcal{C}_2 .

The second code is a rate-1/3 binary linear time-invariant convolutional code \mathcal{C}_2 generated by the degree-1 generators $(D, D, 1 + D)$, i.e., \mathcal{C}_2 is generated by a minimal encoder with

$$\begin{array}{cccc}
\begin{array}{c} 1 + 14x^5 + 25x^6 + 44x^7 \\ x^3 + 2x^4 + 4x^5 + 8x^6 + 16x^7 \\ x^2 + x^3 + 2x^4 + 4x^5 + 8x^6 + 29x^7 \\ x^3 + 2x^4 + 4x^5 + 8x^6 + 16x^7 \end{array} & \begin{array}{c} x^2 + x^3 + 2x^4 + 4x^5 + 8x^6 + 29x^7 \\ x^5 + 3x^6 + 8x^7 \\ x^4 + 2x^5 + 5x^6 + 12x^7 \\ x^5 + 3x^6 + 8x^7 \end{array} & \begin{array}{c} x^3 + 2x^4 + 4x^5 + 8x^6 + 16x^7 \\ x^6 + 4x^7 \\ x^5 + 3x^6 + 8x^7 \\ x^6 + 4x^7 \end{array} & \begin{array}{c} x^3 + 2x^4 + 4x^5 + 8x^6 + 16x^7 \\ x^6 + 4x^7 \\ x^5 + 3x^6 + 8x^7 \\ x^6 + 4x^7 \end{array}
\end{array}$$

Fig. 3. HWAM $\Lambda^{16}(x)$ (modulo x^8) of a section of $N = 16$ time units of Example 1 code \mathcal{C} .

impulse response $(001, 111, 000, \dots)$, whose trellis section is shown in Figure 4(b). The HWAM of this encoder is

$$\Lambda_2(x) = \begin{bmatrix} 1 & x \\ x^3 & x^2 \end{bmatrix}.$$

Since the weights of the $0 \rightarrow 0$ and $1 \rightarrow 1$ transitions are the same for \mathcal{C}_1 and \mathcal{C}_2 , and since the sums of the weights of the $0 \rightarrow 1$ and $1 \rightarrow 0$ transitions are the same, it is evident that the weight distributions of the subcodes $(\mathcal{C}_1)_{[0,N]}$ and $(\mathcal{C}_2)_{[0,N]}$ are the same for all N , and that the free distance spectra of \mathcal{C}_1 and \mathcal{C}_2 are also the same. For the same reason, the weight distributions of the tail-biting terminated codes $(\mathcal{C}_1)_{|[0,N]}$ and $(\mathcal{C}_2)_{|[0,N]}$ are the same for all N .

However, the weight distributions of the projections $(\mathcal{C}_1)_{|[0,N]}$ and $(\mathcal{C}_2)_{|[0,N]}$ are not the same even for $N = 1$. It follows that the weight distributions of the subcodes $(\mathcal{C}_1^\perp)_{[0,N]}$ and $(\mathcal{C}_2^\perp)_{[0,N]}$ of their orthogonal codes \mathcal{C}_1^\perp and \mathcal{C}_2^\perp are not the same, and therefore that their free distance spectra are not the same; this was the point of Shearer and McEliece [8].

On the other hand, since the weight distributions of the tail-biting terminated codes $(\mathcal{C}_1)_{|[0,N]}$ and $(\mathcal{C}_2)_{|[0,N]}$ are the same for all N , it follows that *the weight distributions of the tail-biting terminated codes $(\mathcal{C}_1^\perp)_{|[0,N]}$ and $(\mathcal{C}_2^\perp)_{|[0,N]}$ are the same for all N* .

Since the performance of \mathcal{C}_1^\perp and \mathcal{C}_2^\perp can be analyzed from these weight distributions, it follows that the performance of \mathcal{C}_1^\perp and \mathcal{C}_2^\perp is effectively the same, despite the difference in their free distance spectra.¹ \square

V. CONCLUSION

In summary, similarly to [1], but using tail-biting terminated codes, we have shown that there is a MacWilliams identity between the generating functions of the weight distributions per unit time of a linear convolutional code \mathcal{C} and its orthogonal code \mathcal{C}^\perp , and that this distribution is as useful as the free distance spectrum for estimating code performance. These results effectively resolve the puzzle posed by Shearer and McEliece [8].

ACKNOWLEDGMENT

For an advance copy of [1], I am grateful to R. Johannesson.

REFERENCES

- [1] I. E. Bocharova, F. Hug, R. Johannesson and B. Kudryashov, “A note on convolutional codes: Equivalences, MacWilliams identity, and more,” submitted to *IEEE Trans. Inform. Theory*, 2009.
- [2] G. D. Forney, Jr., “Codes on graphs: Normal realizations,” *IEEE Trans. Inform. Theory*, vol. 47, pp. 520–548, Feb. 2001.
- [3] G. D. Forney, Jr., “MacWilliams identities for codes on graphs,” *Proc. IEEE Inform. Theory Workshop* (Taormina, Sicily), pp. 120–124, Oct. 2009. ArXiv: 0903.4207 [cs.IT].
- [4] G. D. Forney, Jr., “Codes on graphs: MacWilliams identities,” submitted to *IEEE Trans. Inform. Theory*, 2009. ArXiv: 0911.5508 [cs.IT].
- [5] G. D. Forney, Jr., F. R. Kschischang, B. Marcus and S. Tuncel, “Iterative decoding of tail-biting trellises and connections with symbolic dynamics,” in *Codes, Systems and Graphical Models* (B. Marcus and J. Rosenthal, eds.), pp. 239–264. New York: Springer, 2001.
- [6] H. Gluesing-Luerssen and G. Schneider, “On the MacWilliams identity for convolutional codes,” *IEEE Trans. Inform. Theory*, vol. 54, pp. 1536–1550, April 2008. ArXiv: cs/0603013.
- [7] H. Gluesing-Luerssen and G. Schneider, “A MacWilliams identity for convolutional codes: The general case,” *IEEE Trans. Inform. Theory*, vol. 55, pp. 2920–2930, July 2009. ArXiv: 0805.3484v1 [cs.IT].
- [8] J. B. Shearer and R. J. McEliece, “There is no MacWilliams identity for convolutional codes,” *IEEE Trans. Inform. Theory*, vol. 23, pp. 775–776, Nov. 1977.

¹Another way of reaching the same conclusion is to observe that \mathcal{C}_1 and \mathcal{C}_2 are equivalent under a simple time-invariant, finite-memory permutation. Therefore \mathcal{C}_1^\perp and \mathcal{C}_2^\perp are actually equivalent under the same permutation, and thus must have precisely the same performance.